

CLAIMS

What is claimed is:

1. A signal bearing medium tangibly embodying a program of machine-readable
5 instructions executable by a digital processing apparatus to perform a method for data journaling, the method comprising the following operations:
writing data into at least one block in a journal storage area;
marking as dirty, each block in the journal storage area that has been written into;
and
10 determining if the number of blocks in the journal storage area that have been marked as dirty is greater than the number of blocks in a corresponding data storage area that have been used and whose corresponding blocks in the journal storage area are not marked as dirty,
and if so:
15 copying the used blocks in the data storage area whose corresponding blocks in the journal storage area are not marked as dirty, to the corresponding blocks in the journal storage area; and
transmitting to a meta-data server, a message instructing the meta-data server to swap the data storage area and the journal storage area;
20 and if not:
copying the blocks in the journal storage area that have been marked as dirty to corresponding blocks in the data storage area; and
transmitting to the meta-data server, a message instructing the meta-data server to commit the data that has been stored in the journal storage area.
25
2. The signal bearing medium of claim 1, wherein the operations further comprise, transmitting to the meta-data server, for each block in the journal storage area that has been marked as dirty, meta-data indicating that the block has been marked as dirty.

- 30 3. The signal bearing medium of claim 2, wherein:

the operation of marking as dirty comprises using meta-data flags to mark blocks in the journal storage area that have been written into as dirty; and

the operation of transmitting meta-data indicating that the respective block has been marked as dirty comprises transmitting a Harden_Extent_Flags primitive to the meta-data server.

4. The signal bearing medium of claim 1, wherein the operations further comprise, allocating storage for the journal storage area.

10 5. The signal bearing medium of claim 1, wherein the operations further comprise, associating the journal storage area with the corresponding data storage area.

6. The signal bearing medium of claim 1, wherein the operations further comprise, transmitting a message to the meta-data server informing the meta-data server of the association of the journal storage area and the corresponding data storage area.

15 7. The signal bearing medium of claim 1, wherein the operations further comprise: clearing flags in the journal storage area that mark blocks as dirty; and disassociating the journal storage area from the corresponding data storage area.

20 8. The signal bearing medium of claim 1, wherein the operations further comprise, storing in the meta-data server, for each block in the journal storage area that has been marked as dirty, meta-data indicating that the block has been marked as dirty.

25 9. The signal bearing medium of claim 1, wherein the operations further comprise, storing in a local storage/NVRAM, for each block in the journal storage area that has been marked as dirty, meta-data indicating that the block has been marked as dirty.

10. The signal bearing medium of claim 1, wherein the operations further comprise, writing as part of data on a remote storage target, for each block in the journal storage area that has been marked as dirty, meta-data indicating that the block has been marked as dirty.

5

11. The signal bearing medium of claim 1:

wherein the operation of transmitting to a meta-data server, a message instructing the meta-data server to swap the data storage area and the journal storage area, comprises transmitting a Swap_Data_Journal primitive to the meta-data server; and

10

wherein the operations further comprise updating a meta-data cache in a client to reflect that the data storage area and the journal storage area have been swapped.

15

12. The signal bearing medium of claim 1, wherein the operation of transmitting to the meta-data server, a message instructing the meta-data server to commit the data that has been stored in the journal storage area, comprises transmitting a Commit_Data_Journal primitive to the meta-data server.

20

13. The signal bearing medium of claim 1, wherein if in the determining operation it is determined that the number of blocks in the journal storage area that have been marked as dirty is not greater than the number of blocks in the corresponding data storage area that have been used and whose corresponding blocks in the journal storage area are not marked as dirty, then the operations further comprise, serially committing in one transaction, the data in the journal storage area and additional data in a plurality of additional journal storage areas.

25

14. The signal bearing medium of claim 1, wherein the operations further comprise, implementing the following primitives in the meta-data server:

Allocate_Journal_Space;

Associate_Data_Journal;

Swap_Data_Journal;
Commit_Data_Journal;
Disassociate_Journal_Extent; and
Harden_Extent_Flags.

5

15. A computing system, comprising:

a memory; and

a processing device coupled to the memory, wherein the processing device is
programmed to perform operations for data journaling, the operations comprising:

10 writing data into at least one block in a journal storage area;

marking as dirty, each block in the journal storage area that has been written into;

and

determining if the number of blocks in the journal storage area that have been
marked as dirty is greater than the number of blocks in a corresponding data storage area
that have been used and whose corresponding blocks in the journal storage area are not
15 marked as dirty,

and if so:

copying the used blocks in the data storage area whose corresponding blocks in
the journal storage area are not marked as dirty, to the corresponding blocks in the journal
20 storage area; and

transmitting to a meta-data server, a message instructing the meta-data server to
swap the data storage area and the journal storage area;

and if not:

25 copying the blocks in the journal storage area that have been marked as dirty to
corresponding blocks in the data storage area; and

transmitting to the meta-data server, a message instructing the meta-data server to
commit the data that has been stored in the journal storage area.

16. The computing system of claim 15, wherein the operations further comprise:

associating the journal storage area with the corresponding data storage area;
transmitting a message to the meta-data server informing the meta-data server of
the association of the journal storage area and the corresponding data storage area; and
transmitting to the meta-data server, for each block in the journal storage area that
5 has been marked as dirty, meta-data indicating that the block has been marked as dirty.

17. The computing system of claim 15:
further comprising the meta-data server, wherein the meta-data server is coupled
to the processing device; and
10 wherein the operations further comprise storing in the meta-data server, for each
block in the journal storage area that has been marked as dirty, meta-data indicating that
the block has been marked as dirty.

18. The computing system of claim 15:
15 further comprising a local storage/NVRAM coupled to the processing device; and
wherein the operations further comprise storing in the local storage/NVRAM, for
each block in the journal storage area that has been marked as dirty, meta-data indicating
that the block has been marked as dirty.

20 19. The computing system of claim 15:
further comprising a remote storage target coupled to the processing device, and
wherein the operations further comprise writing as part of data on the remote
storage target, for each block in the journal storage area that has been marked as dirty,
meta-data indicating that the block has been marked as dirty.

25 20. A computing system, comprising:
means for writing data into at least one block in a journal storage area;
means for marking as dirty, each block in the journal storage area that has been
written into;

means for transmitting to a meta-data server, for each block in the journal storage area that has been marked as dirty, meta-data indicating that the block has been marked as dirty;

5 means for determining if the number of blocks in the journal storage area that have been marked as dirty is greater than the number of blocks in a corresponding data storage area that have been used and whose corresponding blocks in the journal storage area are not marked as dirty;

10 means for copying the used blocks in the data storage area whose corresponding blocks in the journal storage area are not marked as dirty, to the corresponding blocks in the journal storage area;

means for transmitting to the meta-data server, a message instructing the meta-data server to swap the data storage area and the journal storage area;

means for copying the blocks in the journal storage area that have been marked as dirty to corresponding blocks in the data storage area; and

15 means for transmitting to the meta-data server, a message instructing the meta-data server to commit the data that has been stored in the journal storage area.

21. A method for data journaling, comprising the following operations:

writing data into at least one block in a journal storage area;

20 marking as dirty, each block in the journal storage area that has been written into;

transmitting to a meta-data server, for each block in the journal storage area that has been marked as dirty, meta-data indicating that the block has been marked as dirty; and

25 determining if the number of blocks in the journal storage area that have been marked as dirty is greater than the number of blocks in a corresponding data storage area that have been used and whose corresponding blocks in the journal storage area are not marked as dirty,

and if so:

copying the used blocks in the data storage area whose corresponding blocks in the journal storage area are not marked as dirty, to the corresponding blocks in the journal storage area; and

5 transmitting to the meta-data server, a message instructing the meta-data server to swap the data storage area and the journal storage area;
and if not:

copying the blocks in the journal storage area that have been marked as dirty to corresponding blocks in the data storage area; and

10 transmitting to the meta-data server, a message instructing the meta-data server to commit the data that has been stored in the journal storage area.

22. The method of claim 21, further comprising associating the journal storage area with the corresponding data storage area.

15 23. The method of claim 22, further comprising, allocating storage for the journal storage area.